AN EXPERIMENTAL STUDY OF THE DIFFERENTIATION OF TEMPERAMEMTS ON A BASIS OF RATE AND STRENGTH

By
MILDRED FRANCES BAXTER

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE UNIVERSITY OF MICHIGAN, 1924

January 1927, vol. xxxviii, pp. 59-96

Ann Arbor 1924

AN EXPERIMENTAL STUDY OF THE DIFFERENTIATION OF TEMPERAMEMTS ON A BASIS OF RATE AND STRENGTH

By
MILDRED FRANCES BAXTER

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE UNIVERSITY OF MICHIGAN, 1924

January 1927, vol. xxxviii, pp. 59-96

R. UNIV. BIBLIOTHEEK LEIDEN

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41551527_0052

AN EXPERIMENTAL STUDY OF THE DIFFERENTIATION OF TEMPERAMENTS ON A BASIS OF RATE AND STRENGTH[1]

By Mildred Frances Baxter

I. Introduction 59
II. Experimental Procedure
A. General 60
B. Experiments
(1) For Rate Measures 61
(2) For Strength Measures 69
(3) Summary of Results 72
C. Physiological Measures
(1) Pulse Rate 74
(2) Systolic Blood Pressure 74
(3) Temperature 74
(4) Summary of Results 75
D. Questionaries
(1) Questionary A 77
(2) Questionary B 77
(3) Questionary C 78
(4) Questionary D 79
III. Interrelation of Results 80
IV. Conclusions 95

I. Introduction

As Allport has said in regard to the classification of temperaments, "the four-humor doctrine has haunted the house of psychology for twenty-five centuries."[2] During recent years a new version of this four-fold classification has been put forth. Bahnsen,[3] Wundt,[4] Ribéry,[5] and others have suggested that the four temperaments are a quick-strong, a quick-weak, a slow-strong and a slow-weak type. This classification has repeatedly found a place in text-books as a possible basis for the grouping of individuals with respect to temperament. We have in this research endeavored to determine whether or not this system of differentiation is justified experimentally.

First we shall try to make clear what is meant by our terms quickness or rate and strength. Writers have not agreed on their significance. Bahnsen[6] refers to quick or slow receptivity and to strong or weak spontaneity, Wundt[7] to the quickness and strength of the emotions, and Ribéry[8] to strong and quick motor

[1] From the Psychological Laboratory of the University of Michigan.
[2] G. W. Allport, Personality and character, *Psychol. Bull.*, 18, 1921, 445.
[3] W. Stern, *Die differentielle Psychologie in ihren methodischen Grundlagen*, 1911, 482.
[4] W. Wundt, *Grundzüge der physiologischen Psychologie*, III, 1903, 638.
[5] W. Stern, *op. cit.*, 484.
[6] *Loc. cit.* [7] *Loc. cit.* [8] *Loc. cit.*

and sensory reaction. The terms quickness and rate, as we have used them, require little explanation. The time required for the performance of a task has been taken as the speed index. The situation is very much more difficult when we attempt to determine what has been meant by the term strength and how it shall be defined for the present purpose. There seems to have been no uniformity, and it has been applied to various traits. We learn from Stern[9] that Dorsch speaks of strong or weak bodies and minds, Herbart of strong or weak excitability and manifestation of the emotions; that Hellwig uses these terms in connection with reception and reaction, Kreibig with respect to the will, Klages with regard to motive power and resistance, and Ach with reference to the determining tendency and sensory-motor excitability. According to the interpretation that we have put on the term, strength is that element in efficiency which remains when the speed or rate factor in a performance is disregarded; *i.e.* quality or excellence.

Our aim, then, has been to ascertain whether there appear to be four types of individuals based on combinations of the rate and strength factors, and, in case they do seem to exist, whether they correspond to the types of temperament indicated by descriptions secured through questionaries.

II. Experimental Procedure

A. General

In attacking the problem we first secured 3 sets of scores or measures for both rate and strength. We then endeavored to determine whether any interrelation exists. The first set of scores was obtained from the results of a series of experiments which were selected with a view to including as great a variety of performances as possible. The second set of measures were physiological in character, and included pulse rate as a speed score, systolic blood pressure as a strength score, and temperature. The third set of scores was secured through a group of questionaries. These, in addition to judgments of the rate and efficiency of an *S* in performance, also yielded scores for the speed with which emotions were aroused and for their intensity, of the tendency to be cheerful or to be depressed, for the degree of activity in which *S* usually indulged, and for the tendency to be steady or to fluctuate.

The 138 young women who served as *S*s in this research were university students who were at the time enrolled in the elementary course in psychology. No one of them was acquainted with the purpose of the investigation. Two groups are considered. Group I, composed of 90 *S*s, was used for trying out methods and standardizing materials and procedure, preliminary to the selection of experiments to be used in gathering the final

[9] *Op. cit.*, 481 ff.

data from Group II. The coefficients of reliability, except for the tapping and carbon-copy experiments, and the physiological measures were computed from data obtained from this group. Group II comprised 48 *Ss*, from whom 44 complete sets of data were obtained. A set was considered complete when scores were secured for every experiment, where no more than one of the six questionaries had to be omitted (there were two cases of one omission), and where the physiological measures were secured for at least one morning and one afternoon period. The age range for the members of Group II, with the exception of one *S* aged about 40, was 16 to 25 yrs.; the mode was 19.

The work with Group I extended over a period of 2 years, though that with any one individual was usually completed within 2 weeks. Group II worked for six one-hour periods between March 25 and June 2, 1923. The first and third periods were morning hours, the second and fourth afternoon hours; the last two came in the evening. The first two periods were arranged so that an interval of 28 hr. intervened. It was found to be impossible to carry out this plan quite as rigidly for the third and fourth hours. In 9 of the 44 cases from 2-5 days elapsed between these periods, and in one case 9 days. The evening hours came exactly one week apart.

With Group II the experiments were conducted individually, except for part C of the carbon-copy experiment and for the filling out of questionaries C and D, where the work was done in groups of from nine to fifteen. The procedure was identical for all *Ss*, the order for the first four periods being: (1) experiments in a fixed sequence, (2) physiological measurements, (3) questionary B. Four of the experiments were divided, for the purpose of obtaining sets of scores from which coefficients of reliability might be computed, into two parts which were given on different days. The parts of the carbon-copy, clock problem, and maze experiments were given during the first two periods, those of the tapping experiment during the second and third hours. Where an experiment yields more than one score we shall describe the apparatus or material and general procedure when the experiment is first referred to; the method of obtaining other scores will be treated later.

B. Experiments

(1) *For Rate Measures.* These experiments vary in the degree to which speed was required, ranging from tapping and the learning of nonsense syllables in as few trials as possible, where speed was the only element of interest, through experiments such as the cancellation of words, where the effort was to be divided between speed and accuracy, to those where the rate was purely voluntary. Another aspect in which they are different is the amount of motor activity required in their performance and the degree of mental organization involved. We have attempted to consider these factors in listing and grouping the experiments, placing the emphasis on the first in cases where the two conflicted.

(*R-1*) *Tapping Experiment. Apparatus.* We used a tapping board in the form of an oval brass plate mounted on a 6 × 4 × ⅞ in. block, and a tapping stylus with flexible connecting wire attached. For recording the number of taps an Ewald's chronoscope[10] was placed in the circuit. Before beginning experiments in March we tested the chronoscope by comparing the number of taps recorded by it with the number simultaneously recorded on a kymograph. Light, heavy jarring, and glancing strokes were all tried. The chronoscope was adjusted until the two records, for all but glancing strokes, agreed.

Procedure. Position and method were demonstrated. *S* sat with right arm (except in the case of one who was left-handed) resting comfortably on

[10]Ewald's chronoscope was used without tuning fork as an electric counter.

the table, elbow touching but wrist free. We were careful to see that the stylus was held erect to avoid glancing strokes and was firmly held so that adjustment would not be necessary during the performance. The instructions were: "Tap as rapidly as possible from the signal 'now' to the signal 'stop' which will be given about one half minute later. Pay attention only to your tapping." The switch was opened and stop-watch started as soon as *S* began to tap, and closed at the end of 30 sec. with the giving of the 'stop' signal. Two trials with a one-minute rest interval were given at the beginning of the second and third hours. The average of the first two was correlated with the average of the second two for an index of reliability, and the average of all four trials used for the score. The coefficient of reliability computed from results from 44 members of Group II was found to be .878 ± .0198.[11]

(*R-2*) *Simple Reaction-Time to Sound. Apparatus.* The Hipp chronoscope, a relay, sound hammer, and telegraph key, were set up according to the Dunlap break-to-break system. The initial adjustment of the chronoscope for absolute time and constancy was made with a control pendulum. While the experiment was in progress checks for accuracy were made, at intervals during each day, with a gravity control apparatus which was always in the circuit.

The average m.v. for the test series of 10 times previous to the beginning of an experiment was never more than .6σ and usually much lower, even as low as .0σ. This m.v. was for a time interval of about .483σ and included any error introduced by the control apparatus.

Procedure. We explained the purpose of this experiment to *S* and made her familiar with the stimulus and mode of reacting before the experiment was begun. The interval between the 'ready' signal and stimulus was 2 sec. Thirty reactions were taken. This number was thought to be sufficient for this and for the association reaction-time experiment, on a basis of preliminary work done with Group I. Eighteen series of 50 reaction-times were secured from this group. When the m.v. was calculated for each of these series and for the first half of each series the coefficient of correlation between the m.v.'s for the long and short series was found to be .941 ± .019. Thirty reactions were taken instead of 25, since 30 was the number used by Woodrow[12] in his experiment for the measurement of attention, which we have included in our series of experiments for securing strength scores. Occasionally there were reactions unquestionably premature, and long ones due to lapses of attention, which it was thought should not be included in the average time. The method of eliminating these errors was to omit arbitrarily all reactions below 100 σ and the three longest in every series. The coefficient of reliability, computed from pairs of scores obtained one week apart from 31 *Ss* in Group I, is .589 ± .079.

(*R-3*) *Association Reaction Time Experiment. Apparatus.* An Ach card-exposer apparatus was connected in place of the sound hammer used in the simple reaction-time experiment, and a lip key[13] in place of the telegraph key. A series of 30 familiar words was constructed, including 18 nouns, 6 adjectives, 3 words which might be understood as adjectives or nouns, and 3 which might be nouns or verbs. Only one had as many as 6 letters, and 4 as many as 3 syllables. The words were typed in small letters in the center of white cards 3⅛ × 1⅝ in. Five blank cards were

[11]The formulae used for computing all coefficients and their probable errors are, unless otherwise stated, the familiar product moments formulae.

[12]H. Woodrow, The measurement of attention, *Psychol. Rev. Mon.*, 17, 1914, No. 5.

[13]This key was found to be less satisfactory than was anticipated for it was extremely difficult to get all *Ss* to react in the same manner. Though the instructions were to release the key as the word was spoken, some had a tendency to get the lips away from the key before speaking the word and so were in danger of letting go even before an associate had occurred to them, while others tended to retain their hold on the key while they spoke the word.

introduced at irregular intervals during the series, as a means of detecting any tendency to release the key at the sight of the stimulus word instead of waiting for the appropriate response.

Procedure. *S* was instructed to hold the key closed between her lips until the word exposed had suggested another word but to speak this word as soon as it came to mind. The procedure was demonstrated: 2 practice trials in responding with the first associate were given without the apparatus; then a series of 3 trials using 2 cards with words and a blank one between, was given with the apparatus. Whenever a word in the regular series was reacted to prematurely or while the key was kept closed, it was repeated after several others had intervened. The average of the 30 satisfactory reactions was used as the score. To determine the coefficient of reliability, a second series of words agreeing exactly with the first in regard to the number of nouns, etc., and number of letters and syllables, and similar with respect to meaning, was constructed and given to 31 members of Group I one week after they had had the first list. The coefficient was found to be .698 ± .062.

(*R-4*) *Learning of Nonsense Syllables. Apparatus.* A modification of the Wirth exposure apparatus, arranged by Professor Shepard, was used. The syllables of the series, fastened to a revolving drum, were exposed in succession in an opening 1⅛ × 1⅝ in. A series of 8 nonsense syllables was used;[14] two series were constructed.[15]

Procedure. Series I was used with Group II. The syllables were exposed at the rate of one every 2 sec. *S* was required to spell the syllable aloud during the exposure, as Luh suggested, in order to prevent the forming of meaningful associations. Immediately after every exposure of the series *S* wrote all the syllables that she could recall. The exposures were repeated until the series was written correctly twice with no exposure between these recalls. Frequently one recall would be perfect, but errors introduced into the second reproduction of the list would indicate that the learning was not complete. The rate score was the number of exposures required. Series I and II were learned a week apart by 37 members of Group I. The scores obtained gave a coefficient of reliability for the rate of learning of .663 ± .062.

(*R-5*) *Carbon Copy or 'Accuracy of Movement' Experiment. Part B—Speeded Time. Material.* A complete description of the material will be given here, though the chief purpose for which it was arranged was to get a strength score which will be considered later. The time scores, *R-5*, *R-13*, *R-14*, were secured incidentally and served the purpose of hiding from *S* the real purpose of the experiment. Ten sheets of unglazed onion-skin paper, 8⅛ × 5⅛ in., and 10 sheets of light weight black carbon paper were alternately arranged, the top carbon covered by a sheet of graph paper, and the whole set slipped into a white cardboard frame. The

---

[14]Originally the lists consisted of 12 syllables, but the time required to learn this number was prohibitive.

[15]The series were constructed according to the following plan:
(1) Every syllable was composed of an initial consonant, a vowel, and a final consonant.
(2) No consonant was used twice as the initial or final letter in the same series.
(3) The initial or final consonants of consecutive syllables were not adjacent letters in the alphabet.
(4) No vowel was used more than twice in the same series.
(5) No vowel was used more than once with the same consonant in any of the 16 syllables.
(6) No given vowel followed another given vowel in successive syllables more than once.
(7) No syllable was used that was found to have meaning according to English, Latin, French, or German dictionaries.
(8) Consecutive syllables did not form polysyllabic words.
Rules 1, 2, 6, and 8 are Müller and Schumann's as quoted by E. Meumann (*Psychology of Learning*, trans. by J. W. Baird, 1913, 365 f.); 7 was suggested by C. W. Luh (The conditions of retention, *Psychol. Rev., Mon.*, 31, 1922, No. 142, 3); the others are believed to be new.

right-hand side of the frame was left open when working with individuals, where *E* could prevent handling by *S*; for group work the pads were sealed along that edge also. The graph paper was ruled in light blue, 6 squares to the inch. A black ink margin was drawn just outside the overlapping edge of the frame, and the area within it was divided into sections, one each for *S*'s name, the date and title of the experiment, and three parts, A, B, and C, of the experiment. In each of the last 3 sections three sets of squares, 25 in a row and 2 in height, with a row between sets, were outlined in red. Everything possible was done to keep the apparent but false purpose of the experiment, *i.e.* the measurement of accuracy of movement, before *S*. Uniform No. 3 pencils, kept strictly for this experiment and sharpened after each using, were used by all *S*s in every part of the experiment. In no case so far as is known was the real purpose suspected.

Procedure. The instructions for Part A, *i.e.* for the voluntary time experiments, are given here, since those for part B depend on them. They were: "The purpose of this experiment is to determine how exactly you can control your hand movements. In Part A you are to make three lines of small and large loops, a loop between every pair of blue lines. The small loop is to be one space high, the large loop two spaces high. The line connecting the loops should just touch the lower red line. Be sure to touch the boundaries but do not go beyond them. Begin the first line with a small loop; it will end with a small one, so begin the second line with a large one, and then the third with a small loop. Take your own time; speed does not count." A pattern set of loops was shown to avoid any error. The instructions for Part B, which was given as soon as *S*'s hand was rested, were: "In Part B do exactly the same thing, except that this time do it just as quickly as you can. This time both accuracy and speed count equally. Ready—go." The stop-watch was started with the 'go' signal and stopped as *S* completed the final loop. This time is the score used for speed experiment *R-5*. Group II went through this experiment twice with a 28-hour interval between performances; *i.e.* it was given them during the first and second of their periods. The results for both performances were summed for the scores in all parts of the carbon copy experiment. The coefficient of reliability, computed from the two sets of scores for part B, was .897 ± .0199.

(*R-6*) *Maze Experiments.* It was thought that the kind of performance required in this experiment might bring out differences among the *S*s in their tendency to plunge ahead, sacrificing all else for speed, or to exercise caution.

Material: Two rather simple forms of maze, quite similar to those used with a very different procedure in learning experiments conducted in this laboratory, were employed. Form I was given during the first period and Form II on the second day. The times for the two were averaged for the final score.

Procedure: The instructions were: "I shall give you a maze similar to this one. [A third form was used for demonstration.] When I say 'go, start with your pencil at A (at the point of the arrow) and find your way through; you will come out at B. Keep your pencil on the paper from start to finish. Avoid such blind alleys as this. If you start to go into one I shall say 'no;' when I say 'no,' come out immediately and choose another path. Find your way through as *quickly* and *carefully* as you can. The time you take and the number of wrong turnings count equally." The maze was placed and kept squarely in front of *S*. The stop-watch was started with the 'go' signal and stopped when *S*'s pencil reached B. *S* was required to go through the maze until two successive perfect tracings were made. The time was recorded for all trials, though the first two on each day were the only ones included in the final score, which was an average of these 4 times. The practice-effect, where many trials were necessary, cut

the time to such an extent that the average time in those cases was not comparable with the average time of an *S* whose first two trials were without error. The time-scores for one performance with each of the two forms (one performance = 2 trials) by 80 *S*s from Groups I and II gave a coefficient of reliability of .629 ± .046; the coefficient of reliability, obtained from 31 sets of scores from Group I, where the scores correlated were the total time for the first performance of Forms I and II and the total time of a second performance, a week later, of both forms, was .808 ± .042.

(*R-7*) *Münsterberg Experiment.* This experiment is a repetition, as nearly as could be planned from the descriptions available,[16] of the one which Münsterberg employed in testing the efficiency of electric railway motormen.

Apparatus. A frame held a pack of 12 cards which represented streets. Each of these had lines through the center for car tracks and was scattered with black and red digits which stood for moving objects in the road. A belt, having an opening representing a motorman's window, revolved about the frame parallel to the tracks.

Procedure. *S* was informed regarding the assumed movements of the digits on the cards and was told to see how rapidly she could pass the belt over the card and still detect, through the exposure window, all sections of the track endangered by the moving objects. The space between the tracks was divided into 26 sections, each of which was lettered. In case a moving object and the car reached any section simultaneously its letter was to be called. *S* was warned that accuracy and speed counted equally and was given practice in securing these. *E* slipped each card from under the belt as soon as it had been exposed. The time was taken from the exposure of row A on the first card to the covering of row Z on the last card. This time-score for 30 of Group I for two trials a week apart gave a reliability coefficient of .882 ± .027.

(*R-8*) *Cancellation Experiment. Material.* The passages used for this and other experiments to be mentioned were adapted from a small volume by Dunbar.[17] All three selections as arranged contained 19 lines and 212 words. The *Antwerp*[18] and *Holy Week*[19] passages, which were used for the cancellation, have each 95 words to be cancelled. Those entitled *Caterina* and *Antwerp*, and used for the recall of ideas in the test for logical memory, were each divided arbitrarily into 100 'ideas.' To determine the comparative difficulty, for cancelling, of the first pair of selections, they were given on the same day to 27 members of Group I and again a week later to 23 of these. The resulting data indicated that the two selections were of approximately equal difficulty with respect to the time required and the number of errors made. The *Caterina* and *Antwerp* passages were tested in much the same way and accepted as about equal for reading time and difficulty in the recall of ideas.

Procedure. *S* was required to read the selection aloud, word by word, not by phrases, and cancel every word as she came to it, that had an 'a' or an 'e' or both those letters in it providing there was no 'i' or 'o' in the word. The sense of the passage was to be disregarded. Speed and accuracy were emphasized equally. The purpose of the oral reading was to rule out as far as possible the tendency to reread any parts of the selections which were to be read later for recall. The plan was to use the cancellation performance with the *Holy Week* passage as a control for the cancellation with *Antwerp* where the added task of attending to the sense for the purpose of recall was to be the only changed factor in the experiment. The

[16] H. Münsterberg, *Psychology and Industrial Efficiency*, 1913, 68 ff.; Marx Lobsien, Prüfung der Aufmerksamkeit an Kindern mit der Münsterbergschen Schlittenmethode, *Zsch. f. angew. Psychol.*, 13, 1918, 395 ff.

[17] M. L. Dunbar, *Felicilla's Easter*, 1903, 7 f. [18] *Ibid.*, 13 f. [19] *Ibid.*, 1 f.

time required to complete the task was the score used. The results obtained by giving the *Holy Week* selection twice, a week apart, to 33 members of Group I give a coefficient of reliability of .883 ± .026 for this time. When the time-scores of the first of the two above sets were correlated with the time-scores secured on the same day with the *Antwerp* passage the coefficient is equally high, .884 ± .025, although in the second performance there was introduced a factor not included in the first.

(*R-9*) *Clock Problem Experiment. Apparatus.* Two series, each of 6 clock problems such as Terman [20] uses in his intelligence scale, were used with the exposure apparatus by which the nonsense syllables were shown.

Procedure. The instructions were arranged to suggest the use of visual imagery rather than the seeking for and application of some mechanical means of arriving at the answer. As *S* spoke the answer to the problem exposed the next one was moved into place. The time was taken for the entire series. As previously stated one series was given in each of the first two periods. The rate score is the total of the times required for both series. Two more series were constructed and given a week after the first to 31 members of Group I. The rate scores for the first two and last two series were correlated and gave a coefficient of reliability of .794 ± .045.

(*R-10*) *Mystery Problem Experiment. Material.* The plan of this experiment is similar to a single part of the Burt reasoning test.[21] The mystery story, for which possible solutions have to be selected, is an adaptation from a Scandinavian story by Ingeman.[22] It comprised a single type written sheet. In the final form 14 solutions are suggested on the answer sheet, of which 6 are possible and 8 impossible according to the evidence presented. The first step in standardizing was to seek an arrangement of the material and a procedure that would prove satisfactory from the points of view of scoring and time required. The second step was to compare the results obtained by this procedure with those secured in 3 other tests requiring the thinking through of a situation or problem. At first *S* was required to suggest all possible solutions that occurred to her and signify the one regarded as most probable. This method gave results extremely difficult to score and required too long a time. Next 25 solutions were listed from which *S* selected those she considered possible. In this form the experiment was given twice, a week apart, to 21 members of Group I. We tried various methods of scoring and finally settled temporarily on the following. Only 14 of the 25 solutions were scored. These included the 6 possible ones, 4 which involved one of the 2 impossible conditions, *i.e.* either wounds on the body or the presence of a second person, and 4 which necessitated both of these. Points were given as follows: +1 for every possible solution checked, −1 for every impossible one. The one selected as most probable was given an additional score of from +1 to +6, according to the degree of probability, if a possible solution, and −1 if it involved one of the impossible conditions or −2 if both were required. When the data for the coefficients of reliability for the rate and strength scores had been gathered it was found advisable to make a slight change in the scoring. We gave a score of +4 if no solution involving either of the impossible conditions were checked, −2 if any solution checked involved one of these impossible circumstances, and −4 if both were required by anyone of the solutions checked. To this was added the number of points received for the doubly checked solution. This latter method was used in treating the results on which the coefficients of reliability are based and with the final data; the former method in all other calculations.

---

[20] L. M. Terman, *The Measurement of Intelligence*, 1916, 321.

[21] C. Burt, The development of reasoning in school children, *J. Exper. Ped.*, 5, 1919, 73 ff.

[22] B. S. Ingeman, *The Sealed Room*, 1, 1909, 271 ff.

The tests used for comparison were the Burt Reasoning Test,[23] Syllogism Test D,[24] and the arithmetic problems in Test 5 of the Terman Group Test of Mental Ability, Form B.[25] Two sets of 15 problems from the Burt scale were arranged by the writer and Miss Kurtti to be given by the group method. For the most part only the more difficult problems were used. These were selected to include the widest variety of situations possible, and were divided into 2 balanced sets.

The four tests to be considered were given to 22 members of Group I; the first as an individual test, the others by the group method. Only 10 members of this group could be secured for a second performance of the tests a week later. The coefficients of reliability for the scores were:

Mystery Problem	$.835 \pm .063$
The Burt Reasoning Test	$.685 \pm .115$ (Forms I, II)
Syllogism Test D.	$.850 \pm .058$
Arithmetic Problems	$.798 \pm .075$

To investigate the possibility of a factor common to all 4 tests and to determine to what extent this factor entered into the solving of the mystery problem, every score in the 4 sets was reduced to terms of the standard deviation for the set of which it was a part. This score was expressed as a percentage of the deviation and marked (+) if above the average score and (−) if below it. The 4 reduced scores thus obtained for *S* were averaged for a composite score. The 4 sets of reduced scores were then correlated, using the method described by Woodworth,[26] with this set of composite scores and yielded the following results:

Mystery Problem and Composite Score	$= .712 \pm .067$
Burt Reasoning Test (Form I) and Composite Score	$= .781 \pm .057$
Syllogism Test D and Composite Score	$= .630 \pm .087$
Arithmetic Test and Composite Score	$= .685 \pm .077$

We believe these results indicate the existence of a common factor and justify the use of the Mystery Problem experiment for investigating it.

Procedure. We placed the selection to be read and the answer sheet face down on the table before *S*, the story at her right, then read the following instructions.

"I want to see how well you can think through this problem. It is a mystery to solve. You are to determine how Lieut. Ribot met his death. First read the passage straight through aloud. As soon as you have finished reading it turn this other paper over. There you will find suggested a number of solutions to the mystery. There is one *correct* solution; the others may be possible or impossible. You are to put a check before *every possible* cause of Lieut. Ribot's death and a second check before *the one* which you judge to be the *correct* solution. Before you check the solutions weigh carefully the evidence presented throughout the passage. Part of the evidence points to the one correct solution, part to other possible solutions, and part is misleading, suggesting solutions which *the facts* given show to be impossible. Accept all *statements of fact* as true. You may refer to the passage while checking. Work quickly and carefully. Tell me just as soon as you have finished. Are there any questions? Now begin: read the passage aloud."

We required oral reading to prevent the immediate rereading and studying of sections by some *S*s, and thus insured a more uniform method of attack. The stop-watch was started when *S* began to read the passage and

[23] Burt, *op. cit.*,

[24] *Syllogism Test D*, Dept. Educ. and Psychol., Carnegie Institute of Technology, 1921.

[25] L. M. Terman, *Terman Group Test of Mental Ability*, 1920.

[26] R. S. Woodworth, Combining the results of several tests: A study in statistical method, *Psychol. Rev.*, 19, 1912, 97 ff.

stopped when she reported that she had finished, or, since the instruction to report was frequently forgotten, when she started to hand the paper over or looked up indicating that she was satisfied with the placing of her second check. The coefficient or reliability for the rate score based on 32 cases is .628 ± .072.

(*R-11*) *Voluntary Reading Time.* This rate score was secured as incidental to the strength score *S-9*, and was the rate which *S* voluntarily selected as her usual rate for oral reading.

Material. The *Caterina* selection mentioned under *R-8* was the passage read.

Procedure. For the instructions, see experiment *S-9*. The coefficient of reliability, based on 58 cases, is .852 ± .024.

(*R-12*) *Judgment of Weights, Voluntary Time.* We followed rather closely Whipple's[27] description of Binet's test for depersonalized suggestion by progressive weights.

Materials. A set of 15 weights, of identical size and appearance, and numbered conspicuously from 1 to 15 were used. The first four weighed 42, 63, 84 and 105 grm. respectively; the remaining 11 each weighed 126 grm. These were arranged as numbered in a row, about 2 cm. apart, on a thick black cloth, No. 1, the lightest, at the left, and the eleven heaviest, at the right.

Procedure. The instructions were: "Here is a series of weights, 15 of them. You are to lift them, one after the other, like this, with thumb and fore-finger of the right hand to the height of the weights (10 cm.). As you lift each weight you are to say whether it is heavier, lighter, or the same as the one just before it. Compare each weight as *you lift it.* Here is the first weight, No. 1 at the left of the row." *S* did not know that a record of her rate was being taken; time was not mentioned unless *S* waited for a signal to start; in that case *E* told her to start when ready, that time did not matter. The stop-watch was started when *S* lifted the second weight and stopped when the judgment on No. 15 was spoken. The time was measured from the second weight because some *S*s paused to ask a question while holding the first one. The coefficient of reliability for the rate of judgment, based on 62 pairs of scores, is .711 ± .043.

(*R-13*) *Carbon Copy Experiment. Part A — Voluntary Time.* The material and procedure have been treated in detail under *R-5*. The rate score is the time required for the task. Before this experiment was undertaken at all we endeavored to determine how constant *S*'s 'own time' or 'usual rate' as selected by her would be, and what effect would be exerted on such a rate by one or more speeded experiments preceding it. Twenty-seven members of Group I served as *S*s, though complete records were not obtained from all. Task 1 was to write the numbers 21 through 35, *i.e.* 30 digits; task 2 to write the letters of the alphabet *a* through *z* and again through *d*, *i.e.* 30 letters. Coefficients of reliability were obtained as follows:

(1) Where the 2 performances of task 1, at the voluntary rate, were separated by a short interval, but one in which the same task had been done as rapidly as possible and task 2 both at a voluntary and speeded rate, r = .937 ± .025.

(2) Where the second performance came at the end of the hour, so that a number of tasks, usually at a speeded rate, had been done in the meantime, r = .896 ± .032.

(3) Where the second performance followed the first after an interval of one week, both trials coming at the beginning of the hour, r = .845 ± .05.

(4) Where the first performance came at the beginning of the first hour and the second at the end of the second hour, a week later, r = .796 ± .06.

[27] G. M. Whipple, *Manual of Mental and Physical Tests*, II, 1915, 232.

We thought it safe to conclude that an *S*'s choice of her 'own time' or 'usual rate' was sufficiently constant to be relied upon. The coefficient of reliability for the voluntary rate score, based on the 44 cases in Group II, justified this decision; it was + .813 ± .034.

(*R-14*) *Carbon Copy Rate Ratio.* This ratio, obtained by dividing the speeded time score by the voluntary time score, is used as an index of the degree to which *S* fails to work at the speed of which she is capable, or of the degree to which she sacrifices speed to accuracy. The ratio scores from Group II give a coefficient of reliability of .628 ± .062.

(*2*) *For Strength Measures.* (*S-1*) *Strength-Weight Ratio, (st/wt).* The muscle strength test used to obtain the strength score for this ratio was worked out and partially standardized by Lovett, Martin, and Rich.[28] It measures with a dynamometer the maximum resistance to a steady pull which is sufficient to overcome the contraction of the muscle group under consideration. That this is not a measure of the total potential strength of the muscle tissue, and not, taken alone, an index of physical condition, is evident. Besides the factors of the amount and of the quality of muscle tissue, as Martin and Rich[29] point out, there enters in the factor of the degree to which *S* is able and, it may be added, willing, to innervate the muscles. Rather than the utmost positive effort, the value obtained for each muscle is the 'breaking strength,' the tension shown on the scale at the instant the resistance of the contracted muscle is overcome by a pull in the opposite direction, exerted through the balance. In the complete scale 22 pairs of muscles were measured.[30] Martin and Rich[31] found, however, that a score could be computed for this entire group, from measures secured by testing 4 of the 22 pairs of muscles, which gave a coefficient of correlation of .94 ± .01 with the score gained by actually testing all 44 muscles. The procedure used in this experiment was the abbreviated one by which the strength of the right and left pectoral, forearm flexor, thigh adductor, and thigh abductor muscles was measured.

Apparatus. A light weight flat-face spring balance with a scale capacity of 200 lbs. by 2 lbs. intervals was equipped with a self-registering device. A wooden handle was attached by a swivel to the upper end of the scale and a woven strap attached in a loop about 30 in. in circumference by another swivel to the lower end of the balance. A stout, rather narrow, table with a cleat firmly fastened across one end and a cushion for *S*'s head were necessary for the last 3 pairs of the measurements. A door frame was used for *S* to lean against in the first part, and to aid *S* in steadying herself a rope with a wooden handle was attached to the wall about 10 ft. to the side of *S*.

Procedure. *E* was assisted by an adjustor whose duties were: (1) to place the loop in the assigned position about the arm or leg; (2) to support the loop with one hand and if necessary the arm or leg with the other hand; and (3) to give the commands 'hold back' to mark the beginning of the pull and 'stop' to mark the end. *E* held the handle of the balance in her right hand and the body of the balance in her left. At the command, 'hold back,' *S* contracted the muscle group being tested, and simultaneously *E* pulled the spring-balance, developing tension as rapidly as possible without jerking, and increasing it until the resistance of *S* was overcome. At the 'stop' signal the pull was discontinued, the scale read and recorded. Two trials were given with each group of muscles and the higher of the two measures used. Detailed directions for testing the 4 pairs of muscles

[28] R. W. Lovett and E. G. Martin, Certain aspects of infantile paralysis; with a description of a method of muscle testing, *J. Amer. Med. Ass.*, 66, 1916, 729 ff. Cf. also E. G. Martin, Muscular strength and muscular symmetry in human beings: I. In children, *Amer. J. Physiol.*, 46, 1918, 67 ff.; and E. G. Martin and W. H. Rich, Muscular strength and muscular symmetry in human beings: II. In adult males, *ibid.*, 47, 1918, 29 ff.

[29] *Ibid.*, 39 f. [30] Lovett and Martin, *op. cit.*, 730. [31] *Op. cit.*, 34 ff.

are given by Martin and Rich.[32] The strength of these 8 muscles is equal to 15% of the entire strength by the complete test; therefore to calculate the total strength score, the sum of the measures for these 8 muscles was multiplied by the reciprocal of 15 or 6.67.[33] The ratio of the estimated total strength score to *S*'s weight measure taken at the beginning of this experiment was used as the final score. Martin and Rich[34] considered this a more reliable index, since a coefficient of correlation of .58 ± .04 was found between the strength score and weight. The coefficient of reliability of the ratio, based on 30 cases from Group I, is .749 ± .054.

(*S-2*) *Carbon Copy Score.* The purpose of this experiment is to measure the amount of physical force exerted as pressure during the performance of set tasks. The material and the procedure for part A and B have been taken up in detail under *R-5*. The procedure for part C and the method of scoring will be considered here. Part C was given to the subjects in small groups. The aim here was to set a task difficult to start with and increasing in difficulty to the point where there was little chance of its being done perfectly, so that the *S*s would be obliged to put forth every effort possible. For the first row of this part the instructions were: "This is another test for accuracy. Do it as in Part C. I shall dictate, rapidly, a series of digits which you are to write down as I give them, one digit between every pair of vertical blue lines. The height of the digits does not matter. Leave no spaces; there are 25 spaces and I shall dictate 25 digits. If you miss a digit, guess at it, for I shall not repeat any." *E* illustrated the rate, which was 2 per sec., then read the following series: 3754946281573681492385697. In the instructions for the next section, the *S*s were told, "This time I shall dictate letters rapidly. You are *not* to write the letter I say but the one that *follows* it in the alphabet. As before, write one between every pair of vertical blue lines. The height of the letters does not matter. Leave no spaces; if you are not sure of a letter guess at it." Here the rate was 1 letter per sec.; it was illustrated before the following series was read: aplgwcxnitfgyhobsjmudkvwe. The instructions for the last part were: "This time you are to write numbers. I shall give a number of two digits and you are to write the number which you get by subtracting one from the first digit and adding one to the second digit. Put one digit in each space, *i.e.* use 2 spaces for each number. Leave no spaces." To make sure the directions were understood the method was demonstrated on the blackboard with the number 58. This series was read at the rate of one every 2 sec. The numbers used were 72, 46, 60, 38, 81, 47, 28, 32, 97, 53, 25, 64.

The method of scoring was to count, for every part, including the heading where *S* wrote her name, the date, and Accuracy of Movement Test, the number of sheets of onion skin paper on which any of the carbon could be detected. Care was taken to do this in a bright but not glaring light. At times it was necessary to examine the carbon to differentiate between markings in the paper and carbon on its surface. If there was carbon on the second sheet under the heading, on the third under A, on the fourth under B and on the sixth under C, the score would be 15. As has been said, this experiment was given twice to Group II. The sum of both scores was used for the final strength score. The coefficient of reliability computed from these two sets, is .833 ± .030.

(*S-3*) *Woodrow's Measurement of Attention Experiment.*[35] This experiment may measure attention as Woodrow claims, or it may get at a more highly organized mental process, that of holding things (signal, stimulus,

---

[32] *Op. cit.*, 35 f.

[33] L. M. Moore and J. L. Barker, Monthly variations in muscular efficiency of women, *Amer. J. Physiol.*, 64, 1923, 3, 405. In their work, these investigators multiplied the sum of the 8 measures by 7.04, since this sum has been found to be 14.2% of the total strength in the case of women where it is 15% for men.

[34] *Op. cit.*, 38 ff. [35] *Op. cit.*

reaction, etc.) in order, and measure *S*'s ability to carry out a type of plan, as Professor Shepard suggests. Which supposition is nearer the truth does not matter here, since we were satisfied that the experiment did give a measure of something not touched by the other experiments.

Apparatus and Procedure. This was just as in the simple reaction-time experiment which preceded it and furnished one of the 2 scores used in computing the final score. The instructions were: "In the next part of the experiment the interval between the 'ready' signal and the stimulus will vary in length. Otherwise it is identical with the first part. React *as quickly as you can*, just as soon as you hear the stimulus. Do your very best *every reaction.*" The set of irregular intervals used was the one given by Woodrow.[36] Practice was given on the last 3 of the set before the times were recorded. The score is the absolute difference between the average reaction time for all the 2 sec. intervals and the average for all the irregular intervals. A low score is taken as an indication of strength. The coefficient of reliability for this score, based on 2 sets of 31 scores secured a week apart, was .425 ± .099.

(*S-4*) *The Maze Experiment.* The material and procedure have been given in detail under *R-6*. The strength score used was the total number of errors made in all trials with both forms. A coefficient of reliability of .402 ± .05 was obtained when 2 sets of 80 scores each were correlated. The scores in one set were the total number of errors made in all trials with Form I on the first day that it was given; those in the other set the total number made with Form II on the first day that it was given. Thirty-one *S*s worked with both forms of the maze again a week after their first performance with them. The coefficient found between the total number of errors made with Form I and II the first week, and the total number made the second week was .429 ± .099.

(*S-5*) *Münsterberg Strength Score.* Here again it is uncertain just what process or processes were being tested. It is probable that breadth of vision, perception, and attention—either its breadth or its flexibility—played a large part in the performance, which required the noting of the color, form, and locality of all the digits. The ability to construct speedily a systematic procedure quite probably increased efficiency to a great extent. The apparatus and procedure have been given in detail under *R-7*. The score used here was the total number of danger points that were omitted; though recorded, no attention was paid to the letters wrongly called. The coefficient of reliability, computed from 2 sets of 30 scores obtained a week apart, is .869 ± .031.

(*S-6*) *Cancellation Experiment.* An account of the material and procedure for this experiment has been given under *R-8*. The score used here was the total number of errors, both words wrongly cancelled and those omitted. The coefficients of reliability for this score, when computed for 2 trials a week apart, both using the *Holy Week* selection, with 33 *S*s, is .800 ± .042. When based on the first trial by 35 *S*s with *Holy Week* and the first trial with *Antwerp*, given on the same day, the coefficient of reliability is .875 ± .027, though a new factor, that of attending to the sense of the passage for the purpose of recall, entered into the latter task.

(*S-7*) *Clock Problem Experiment.* The material and procedure have been given under *R-9*. The score was the total number of problems in both series correctly answered. An answer was considered correct if it fell within the five minute interval; for instance, should the answer be 3:23, any answer between 3:20 and 3:25 inclusive would be accepted as correct. To determine the coefficient of reliability a third and a fourth set of 6 problems each were given, a week after the first sets, to 31 *S*s; it was found to be .638 ± .072.

[36] *Ibid.*, 127.

(*S-8*) *Mystery Problem Experiment.* The material, procedure and method of obtaining the strength score have been described under *R-10*. The experiment was performed a second time, a week later, by 32 *S*s. They were instructed to go at the problem as carefully as though it were for the first time; they might arrive at the same conclusions or different ones. They seemed to carry out the directions quite faithfully, and many changes occurred in the second series of checkings, though the same types of solutions were considered possible. These data gave a coefficient of reliability of .502 ± .089.

(*S-9*) *Logical Memory Experiment. The Recall of Ideas. Material.* The *Caterina* selection mentioned under *R-8* was used.

Procedure. The instructions were, "This experiment is one to see how much of a selection you can recall after reading it through once. When I say 'go' turn the paper over and read the selection straight through aloud. Read it at your usual rate for oral reading. Put your attention on the sense rather than on your manner of reading. As soon as you have finished reading turn the paper over, and write as much of the passage as you can recall. Use the words of the passage as far as you can remember the exact wording; when you do not recall it, use your own words. Read it straight through once. Go." After *S* finished writing she was further directed to put down any other separate words or phrases recalled out of context.[37] An idea recorded was considered satisfactory if it approximated the original to a rather close degree. The experiment by Henderson[38] served as a guide in this respect. The score was the total number of ideas recalled. When the *Caterina* passage was given twice, a week apart, to 33 *S*s, the scores gave a coefficient of reliability of .764 ± .049; when the scores for *Caterina* and for *Antwerp* were correlated, though the latter included a factor not in the former, *i.e.* cancelling, as with the *Holy Week* selection, while it was being read for recall, the coefficient of reliability for recall for 35 cases was .782 ± .044.

(*S-10*) *Suggestibility Experiment.* The material has been given under *R-12*. The score was the number of weights that were incorrectly judged as heavier. Here as in *S-3* a low score is considered indicative of strength. The coefficient of reliability for scores secured a week apart from 62 *S*s is .567 ± .058.

We had planned for two other strength scores, but gave them up because of their apparent unreliability. The score for the written recall of the nonsense syllables after a lapse of 28 hours gave a coefficient of reliability of .105 ± .12. We had hoped to secure an index of the breadth or the flexibility of attention by having *S* cancel words in the *Antwerp* selection, as she had done with the *Holy Week* passage, *while* she read it straight through for recall. The ratios of the cancellation time, cancellation errors and recall scores in the single tasks, to the same scores in the double task were to have been combined for this final score. Although the lowest coefficients of reliability for any one of the 6 individual scores was .659 ± .066, the coefficients for the ratios were disappointing. The time ratio gave a coefficient of reliability of .595 ± .076, the error ratio a coefficient of −.018 ± .114, and the recall ratio one of .247 ± .110. It was evident that a score composed of ratios apparently so unreliable would itself have slight if any reliability.

(*3*) *Summary of Results.* We have expressed the individual scores in terms of rank, 1 signifying the highest place for rate

[37]She was also requested to underline every word that she was certain had occurred in the passage, but this score was not used, since we thought a low score might in some cases indicate weakness, as shown in lack of self-confidence, and in other instances strength, as evidenced by caution.

[38]E. N. Henderson, A study of memory for connected trains of thought, *Psychol. Rev. Mon.*, 5, 1903, 34.

and strength, 44 the lowest place. See Table I and II for sample cases.[39] Composite scores for every *S* for rate and strength were computed in two ways. One composite score was obtained by averaging the ranks secured in all tests of the rate series and again all in the strength series. For the other score the number of times that *S* gained a place in the quicker half of the group was expressed as a percentage of the number of possibilities of being in the group considered, *i. e.* as a percentage of the number of experiments in the series. The two forms of composite rate scores gave a coefficient of correlation of .894 ±.024, and those for the strength scores one of .859 ±.029. Because of the close

TABLE I
Rankings in Rate Experiments

S	R-1 Tap.	R-2 S.R.T.	Av. R. 1 & 2	R-3 A.R.T	R-4 N.Syl.	R-5 CC.S	R-6 Maze	R-7 Mun.	R-8 Can.	Av. 3,4, 5, 6, 7, & 8
A	12.0	22.5	17.5	9.0	3.0	9.0	35.5	29.0	13.0	21.6
B	40.0	25.5	32.8	38.0	8.0	42.0	43.0	27.5	25.0	34.4
F	42.0	27.0	34.5	43.0	32.0	32.0	33.0	12.0	42.0	29.8
M	41.0	20.5	30.8	12.0	14.0	19.0	41.0	11.0	4.5	18.9

S	Av. R 7 & 8	R-9 Cl. P.	R-10 Myst.	R-11 R.T.	R-12 J.Wt.	R-13 CC.V.	Av. 11 12 & 13	R-14 CC/SV	Com. Score % Meth	Com. Score Av.R.
A	21.0	33.0	38.0	16.0	1.0	10.0	9.0	15.5	68	17.6
B	26.3	3.0	33.0	39.5	43.0	29.0	37.2	7.5	21	28.9
F	27.0	42.0	22.0	41.0	17.0	40.0	32.6	35.0	21	32.8
M	7.8	13.0	17.0	19.0	11.0	14.0	14.7	13.0	86	17.9

TABLE II
Rankings in Strength Experiments

S	S-1 S/W.	S-2 CC.	S-3 W's A.	S-4 Maze	S-5 Mun.	S-6 Can.	Av.R. 4 % 5
A	10.0	18.0	36.0	16.5	22.0	6.0	14.8
B	1.0	15.5	42.0	23.0	14.0	35.0	24.0
F	10.0	24.0	16.0	3.5	38.0	35.0	25.5
M	3.0	11.0	43.0	16.5	38.0	38.0	30.8

S	S-7 Cl P.	S-8 Myst.	S-9 R.Id.	Av. R 7, 8, 9	S-10 J. W.	C. Sc. %	C. Sc. Av.R.
A	39.0	16.5	7.0	20.8	14.5	80	18.6
B	15.5	16.5	1.0	11.0	6.5	75	17.0
F	31.5	25.0	33.0	29.9	27.5	35	24.4
M	39.0	33.5	21.5	31.3	41.0	35	28.5

[39] Copies of Tables I, II, IV, and VII in their complete form are on file in the University of Michigan Library and may be borrowed under the usual courtesies of library exchange.

correlation only one, the percentage score, has been considered to any extent. Further presentation of the results and their discussion will be considered later.

C. Physiological Measures

Temperature, pulse rate, and blood pressure were taken at the end of the period with *S* seated. It was seldom necessary for *S* even to leave her chair more than once during the hour, so that the measures were taken after a comparatively long period entirely free from any muscular activity which would be likely to influence them. All *S*s had spent the time in performing the same mental tasks. They were comfortably seated beside the table on which they had been working and were instructed to relax, to rest, and "think about nothing in particular." The temperature was allowed to register in the mouth while *E* counted the pulse rate; blood pressure was taken last.

(*1*) *Pulse Rate.* *E* counted the pulse at the left wrist for two or three 30-sec. intervals to make sure that it was constant, and that the first rate had not been influenced by the newness of the situation. The rate for 30 sec. was doubled for the measure. The final rate used was the average of the 4 obtained during that number of periods. For the coefficient of reliability see Table III, and for samples of the measures, Table IV.

Table III
Coefficients of Reliability for Physiological Measures

Measured during	Pulse Rate	Blood Pressure	Temperature
A.M.	.691±.056	.679±.058	.212±.101
P.M.	.734±.049	.638±.063	.523±.076
A.M. + P.M.	.792±.041	.716±.054	.492±.084

Table IV
Physiological Measures

	Pulse Rate					Blood Pressure					Temperature				
S	AM$_1$	PM$_1$	AM$_2$	PM$_2$	Av.	AM$_1$	PM$_1$	AM$_2$	PM$_2$	Av.	AM$_1$	PM$_1$	AM$_2$	PM$_2$	Av.
A	84	84	74	74	79	104	96	96	92	97	97.2	98.6	98.2	98.4	98.3
B	80	76	76	60	73	92	90	96	96	94	98.2	99.0	98.4	98.6	98.6
F	96	96	92	86	91	100	88	80	84	88	99.0	99.0	97.6	98.0	98.4
M	74	76	76	68	74	110	102	104	106	106	96.6	97.4	97.6	97.4	97.3

(*2*) *Systolic Blood Pressure.* The Tycos sphygmomanometer was used with the auscultatory method. The compression sleeve was placed about the bare left arm just above the elbow, and the stethoscope applied over the radial artery at or above the elbow. Except in one or two instances where *S* was restless because of discomfort or haste, 2 or 3 readings were taken at each period, with rest intervals of at least a minute, for verification and to eliminate the possibility of the first rate's being high because of *S*'s unfamiliarity with the procedure. The measure used was the average of the 4 measures secured on 4 different days. For the coefficient of reliability see Table III, and for samples of the measures, Table IV.

(*3*) *Temperature.* A Tycos fever thermometer, requiring one minute for registration, was used. More than the required time was allowed and the reading taken at least twice to make sure that the temperature was not still rising. See Table III for the coefficients of reliability and Table IV for a few of the readings.

(4) *Summary of Results.* It will be noted (see Table IV) that the average temperature and blood pressure were low. Burlage[40] in a study of 1700 cases of public school and University girls found the average pulse rate for girls of 18 years to be 79.78 and to fluctuate little up to 26 years of age. Our average agrees closely with this. Schneider and Truesdell[41] found the average blood pressure of the group of normal women whom they were studying to fluctuate from 95 to 107 mm., with a midpoint of 101, for recumbency; and from 104 to 113.7 with a midpoint of 108.7, for standing position. Alvarez,[42] studying pressures obtained from 5807 women between the ages of 16 and 40 by palpation with *S* in standing position,[43] found the average for the entire age range to be 115 mm. This however, is higher than would be expected for our group, since he also found that, instead of the average pressure rising steadily with age as it had been supposed to do, it dropped between the ages of 17 and 25 years, which are the age limits that include all but 2 of our cases. The range for women proved to be practically 85 to 155; there were 4 cases, ages 16 to 21, with pressures between 80 and 84. We believe that our measures may be slightly low because taken after an hour of comparatively slight body activity but rather strenuous mental exertion. O'Hare,[44] in a study of normal and vascular hypertension cases, found that physical rest caused a marked fall in blood pressure, the range being from 0 to 46 mm. with an average of 21 mm. Pillsbury and Griffitts,[45] found that pressure also fell during a period devoted to the performance of mental tasks. It is possible also that the circumstance that the measures were all taken in the spring, during the last two months of the school year, was a contributing factor.

Since authorities do not seem to agree[46] on the presence or absence of a diurnal rhythm in blood pressure and pulse rate,

[40]S. R. Burlage, The blood pressures and heart rate, in girls, during adolescence: A statistical study of 1700 cases, *Amer. J. Physiol.*, 64, 1923, 259.

[41]E. C. Schneider and D. Truesdell, Daily variations in cardiovascular conditions and a physical efficiency rating, *Amer. J. Physiol.*, 67, 1923, 199.

[42]W. C. Alvarez, Blood pressure in university freshmen and office patients, *Arch. Int. Med.*, 26, 1920, 381 ff.

[43]The measures in this study would tend to be higher than ours because of the difference in *S*'s position, and on the other hand would tend to be lower because of the difference in the method used. These tendencies would probably just about compensate, making the data comparable with ours.

[44]J. P. O'Hare, Vascular reactions in vascular hypertension, *Amer. J. Med. Sci.*, 159, 1920, 370.

[45]C. H. Griffitts and W. B. Pillsbury, An experiment on indirect measures of fatigue, *Proc. VII. Inter. Cong. Psychol.*, 1924, 212 ff.

[46]Schneider and Truesdell, *op. cit.*, 194 ff. These investigators state that early students of pulse rate frequency came to the conclusion that it "is greater in the morning than in the evening," and that the heart was more excitable in the morning; also that Weysse and Lutz "concluded that the systolic pressure rises an average of 8 mm. immediately after the ingestion of food, then gradually falls until the beginning of the next meal, and that there is a general rise during the day, whereas their observations lend support to the more general increase in rate throughout the day and fail to indicate a decided diurnal change in the excitability of the heart," and show no general day-time rise in the level of arterial pressures for the groups that ate the usual meals, though such a gradual rise was found in a group that went without meals.

we averaged all measures taken at each hour from 9 A. M. to 5 P. M. The results are stated in Table V. The differences between the averages for measures taken at different hours were less than the variations found in many instances in measures taken for the same *S* at the same hour on two days. We concluded that it would therefore not be safe to associate unexpected findings in any case with the hour factor.

TABLE V
Distribution of Physical Measures by Hours

Hr.	Pulse Rate		Blood Press.		Temperature	
9	80 (*23*)		100 (*23*)		97.8 (*23*)	
	=79		=100		=98.0	
10	78 (*16*)		100 (*16*)		98.2 (*18*)	
		=79		=99		=98.2
11	80 (*25*)		98 (*26*)		98.1 (*26*)	
	=78		= 98		=98.3	
12	75 (*21*)		98 (*20*)		98.5 (21)	
2	83 (*24*)		100 (*24*)		98.1 (*22*)	
	=80		= 98		=98.2	
3	77 (*19*)		95 (*19*)		98.3 (*19*)	
		=79		=98		=98.3
4	81 (*21*)		95 (*21*)		98.4 (*21*)	
	=78		=97		=98.4	
5	74 (*20*)		98 (*20*)		98.4 (*20*)	

The figures in *italics* give the number of measures.

Another point on which there is disagreement[47] is that of the presence, (and nature, if present), or the absence of a monthly rhythm in cardio-vascular activities in women, and the presence or absence of a periodic mental and motor inefficiency. To determine whether or not this factor should be considered in our study, the average of all measures for every one of the 3 physiological processes was found for the menstrual period, for the 5 days previous to the onset, for the 5 days following the close, and for the inter-menstrual period. Our results, stated in Table VI, show slight differences for the 4 periods and are in

[47] L. M. Moore and C. R. Cooper, Monthly variations in cardio-vascular activities and in respiratory rate in women, *Amer. J. Physiol.*, 64, 1923, 416 ff. According to these writers the theory of presence of a rhythm was put forward by Jacobi in 1876 and held sway until investigations during the last decade in general produced evidence in apparent refutation. Their own study (*ibid.*, 420), however, and that by King (J. L. King, Concerning the periodic cardio-vascular and temperature variations in women, *Amer. J. Physiol.*, 34, 1914, 203 ff.) both support the rhythm theory, though they do not seem to agree concerning the order of the phases. King, as did Jacobi and others, found that there was "rhythmical movement in the life-processes," the highest point of which was reached from 2 to 4 days before the menses, the lowest point about 3 days after their cessation; whereas Moore and Cooper (*op. cit.*, 423) found that the typical variations are a low menstrual and a high inter-menstrual period. An investigation by Moore and Barker (L. M. Moore and J. L. Barker, Monthly variations in muscular efficiency in women, *Amer. J. Physiol.*, 64, 1923, 414) suggests a "menstrual decrease in strength." Another investigation, one by Hollingworth (L. S. Hollingworth, *Functional Periodicity*, 1914, 94) gives evidence in favor of there being no "periodic mental and motor inefficiency in normal women."

general agreement with the conclusion reached by King,[48] and by Jacobi and other early workers mentioned by King, rather than with the findings of Moore and Cooper.[49] However, when

TABLE VI

Distribution of Psychological Measures (averages) with Respect to Menstruation

	Pulse Rate	Blood Press.	Temp.
Intermenstrual period	76 (*77*)	97 (*77*)	98.0 (*78*)
Within 5 days of onset	82 (*33*)	100 (*33*)	98.5 (*33*)
Menstrual period	82 (*32*)	101 (*32*)	98.3 (*32*)
Within 5 days after close	77 (*20*)	97 (*20*)	98.0 (*20*)

Figures in *italics* gives the number of measures.

the records of our individual *S*s were studied, the directions of the variations concomitant with differences in physical condition were not consistent. Also similar differences were found between measures for the same *S*, all taken during the same type of period; and, on the other hand, identical measures were obtained for an individual at different phases of the cycle. We therefore concluded, as in the case of the distribution of measures by hours, that it would not be safe to associate unexpected findings with this periodic factor.

The record of physical examinations and medical attention received while attending the university was investigated for all *S*s, with special note of any reference to thyroid abnormality. Such record as there was did not indicate, however, the nature of the disturbance, so that it was impossible to tell whether one should expect pathological cardiovascular conditions, and in which direction any change which might be present had occurred.

D. QUESTIONARIES

(*1*) *Questionary A.* Questionary A was for the purpose of obtaining general information and of giving instructions. It stated that all appointments must be kept in order to secure credit for the work, that the scores in the experiments would not affect the grade in the course provided all parts were conscientiously done, and that they should not discuss the experiment with any one taking the elementary course.

(*2*) *Questionary B.* A daily report of *S*'s physical and mental condition was secured through this questionary, which was filled out at the close of each of the first four periods. She was asked to report concerning the number of hours she had slept the

[48] *Op. cit.*, 218; cf. footnote 47. [49] *Loc. cit.*

previous night, whether sleep was continuous or broken, to note the affective and emotional quality of dreams, to state whether she had eaten breakfast and lunch on that day, and to evaluate her physical condition and 'frame of mind' according to a scale given her.

(3) *Questionary C.* This questionary was filled out twice, a week apart, by every *S* for herself. It was then filled out by a member of *S*'s immediate family, preferably her mother, and by three young women friends. (Through misunderstanding 6 of the 132 in the last group were filled out by men friends; this was not discovered in time to correct the error.) The intra-group method of procuring judgments could not be used, since the *S*s were for the most part strangers to each other. The following instructions were printed on the first sheet of the questionary:

"Please fill out this questionary carefully and frankly. It will be considered absolutely confidential and used only in connection with the research problem of which it is a part.

A number of characteristics have been arranged in pairs: 1, (a) and (b); 2, (a) and (b), etc. Each pair is composed of two opposed traits. To fill out the questionary you are to assign a group number to each characteristic (a) and to each characteristic (b). First imagine all the young women *of the population at large* who are between the ages of 17 and 25 years and who have had at least average educational advantages, that is, at least a year or two of high school work, divided into six groups according to the degree to which they possess each of the characteristics (a) and each of the characteristics (b). Those in group 1 are the one sixth of the entire section of the population selected for comparison (chosen on a basis of age and education) who have the characteristic under consideration, for instance 1 (a), to the least degree. Those in group 2 are the sixth having it to a greater degree and so on. Group 6 will be the sixth who have it to the greatest degree. The same will be true when trait 1 (b) is considered, except that the grouping will be reversed, since the traits are arranged in pairs of opposed characteristics; those who were in group 1 on a basis of trait 1 (a) will be in group 6 on a basis of trait 1 (b) and those in group 4 for characteristic 1 (a) will be in group 3 for 1 (b). To assign a group number to the characteristics listed below, decide in which of the six groups the person for whom you are filling out the questionary would belong with respect to each of the traits, first, 1 (a), and then 1 (b), and place the numbers of the groups in the appropriate places; next do the same for 2 (a), then for 2 (b) and so on. *Do not omit any.* The *sum* of the two group numbers assigned for any pair *must be seven.* Do not place the person between groups. You must place her in one of the six groups. Read the entire list through once before you begin to assign group numbers. Be careful to distinguish between characteristics which at first appear to be identical; no two are just the same."

A few of the sixty traits are listed here:

1. (a) To be quick in her movements,
   (b) to be slow in them. (a)——(b)——
4. (a) To work carefully and accurately,
   (b) to work carelessly and inaccurately. (a)——(b)——
18. (a) To be fluctuating,
   (b) to be resolute. (a)——(b)——
26. (a) To be gloomy and depressed,
   (b) to be cheerful and happy. (a)——(b)——

31. (a) To become angry easily, to be quick tempered,
(b) to be slow to anger and only under great stress. (a)——(b)——
47. (a) To act after deliberation,
(b) to act impulsively. (a)——(b)——

The method of scoring was to consider separately the traits, *i.e.* either (a) or (b), which were thought to indicate quickness, strength, activity, steadiness, or a tendency to be cheerful. We shall speak of these as positive traits. There are 14 traits considered indicative of quickness, 29 which were thought to be suggestive of strength, 6 were taken as indicative of activity and 7 to indicate steadiness. Seven characteristics were thought to point to a tendency to be cheerful; ease of, or quick initial arousal of, the emotions seemed to be suggested by 6; and an idea of the intensity or strength of the emotions may be gained from 4 traits.

For the final score for a trait the 2 numbers assigned by *S* to a trait were averaged and this score averaged with the 4 numbers assigned by the other persons to the same trait. With the scale of rating used, the greatest possible range for the 44 scores for any one trait is 5. The least actual range for any one of the 60 traits was 1.9, the maximum 5, and the average range 3.5. To find the composite rate and strength scores for the questionary, we found the average score for each of the positive traits, then gave as a rate score for *S* the percentage of times, out of the 14 possibilities, that she gained a score above the average for a trait in the rate group, and as a strength score the percentage of times she secured a rating above average in the 29 strength traits. The composite scores for the remaining groups, where the number of traits included was smaller, were the total number of points gained in a group by receiving a score of +1 for each trait within the group for which *S* secured a rating above the average for the trait and −1 for each trait in which she was below average. A few of these scores are given in Table VII.

TABLE VII
Questionary C Scores

S	Rate	Strength	Steadiness	Activity				Emotions		Mood	
				Physical	Mental	Both P&M	Total	Arousal Rate	Intensity	Score	Steadiness
A	89	69	+1	0	+3	+1	+4	+6	+4	+4	−1
B	89	55	+1	0	+3	+1	+4	+4	0	−1	−1
F	7	48	+3	−2	−1	−1	−4	−4	0	+4	−1
M	64	86	+5	0	−1	+1	0	−4	+2	0	+1

(4) *Questionary D.* *S* filled out this questionary after completing Questionary C. Part A consisted of a list of 35 activities such as baseball, dramatics, designing or drawing, and cooking. *S* was instructed to consider herself a participant in these and assign a value of −1 if displeasure or unpleasantness was derived from an activity, of 0 if neither displeasure nor enjoyment was derived, and a value of 1, 2, 3, or 4 as the degree of pleasure increased. Part B inquired into *S*'s habits of sleep; information concerning past and present health conditions was sought in parts C, D, and F. The questions in part E pertained to nervous habits such as the biting of the finger-nails and rocking. We made use of these data in the case studies which follow the statistical treatment of the results obtained from the experiments and the preceding set of questions.

III. Interrelation of Results

In Tables VIII and IX will be found the coefficients of correlation between the various experimental scores and between each score and the composite score for the series. The second composite rate score used in Table VIII was computed on a basis of the scores which had both a coefficient of reliability of .700[50] or more and a coefficient of correlation with the first composite score of at least .270. The second composite strength score in Table IX was based on the scores which had a coefficient of reliability of .429 or higher and a coefficient of correlation with the first composite score of at least .292. The coefficient of correlation between the two composite rate scores and between the two composite strength scores, however, are .906 ±.018 and .908 ±.009 respectively; also the coefficients of correlation between each one of the selected experiments and the second composite score were not very different from those obtained for the same experiment with the first composite scores. The 8 rate experiments gave an average coefficient of correlation of .426 with the first composite score and of .448 with the second composite score; the 7 strength experiments gave an average coefficient of correlation of .418 with the first composite score and of .454 with the second composite score. For these reasons we have used, to any extent, only the composite scores based on the entire series of experiments.

The coefficients of correlation between the 14 rate experiments can scarcely be said to give evidence of any one factor running through all. Fifty-eight of the 91 coefficients for pairs of experiments are positive; only 4 are at least 5 times their P.E., and 2 are 4 times it. The coefficients between the separate scores and the composite score are higher for the most part; 13 of the 14 are positive, 5 of these are 5 times the P.E. and one is 4 times it. The correlations are highest between the experiments in the group where the time was purely voluntary. The Münsterberg and cancellation experiments where speed and accuracy were stressed equally correlate relatively well with each other, and with the voluntary times for reading a selection and for judging the weights. The experiments requiring maximum speed do not here correlate well either with each other or with those of the other two types just mentioned. A similar impression is gained from an inspection of the coefficients for the strength scores. Of the 45 coefficients of correlation be-

[50] Probable errors for coefficients of correlation where N = 44 are:

r	P.E.	r	P.E.	r	P.E.	r	P.E.
.000	.102	.250	.095	.500	.076	.750	.044
.050	.101	.300	.093	.550	.071	.800	.037
.100	.101	.350	.089	.600	.065	.850	.028
.150	.099	.400	.085	.650	.059	.900	.019
.200	.098	.450	.081	.700	.052	.950	.009

TABLE VIII

Coefficients of Correlation between Rate Scores

	R−1 Tap	R−2 S.R.T	R−3 A.R.T	R−4 N.Syls	R−5 CC.S	R−6 Maze	R−7 Mun.	R−8 Can.	R−9 Cl.P.	R−10 Myst.	R−11 R.T	R−12 J.Wt.	R−13 CC.V	R−14 CC S/V	Pulse Rate	Temp.
R−1	+.*878*															
R−2	+.236	+.*589*														
R−3	+.022	−.020	+.*698*													
R−4	−.242	−.094	−.198	+.*663*												
R−5	+.059	−.211	+.098	+.096	+.*897*											
R−6	+.175	−.126	+.018	−.204	+.224	+.*809*										
R−7	+.112	+.071	+.010	+.212	−.020	+.196	+.*882*									
R−8	+.229	+.103	+.039	+.303	−.091	+.115	+.495	+.*883*								
R−9	+.120	+.041	+.329	+.150	−.041	+.246	+.181	+.454	+.*794*							
R−10	−.185	−.221	−.099	−.071	+.144	+.004	+.003	−.118	+.021	+.*628*						
R−11	+.327	−.038	−.238	+.181	+.121	+.055	+.305	+.385	−.033	+.024	+.*852*					
R−12	−.006	−.231	+.224	+.213	+.279	+.057	+.358	+.314	−.026	−.043	+.284	+.*711*				
R−13	+.041	−.130	+.016	+.031	+.715	−.037	−.053	−.025	−.153	+.273	+.297	+.236	+.*813*			
R−14	−.187	−.006	+.033	+.176	−.113	−.270	−.140	+.047	−.140	+.234	+.212	+.069	+.558	+.*628*		
P.R.	+.127	−.162	+.157	−.452	−.032	−.189	+.066	−.067	−.300	+.059	−.055	+.273	−.075	−.143	+.*792*	
Comp. Sc. 1.	+.094	−.03	+.239	+.293	+.277	+.271	+.419	+.587	+.366	+.144	+.455	+.581	+.455	+.275	+.098	−.0096
Comp. Sc. 2.					+.405	+.340	+.418	+.571	+.316		+.441	+.621	+.469			
Av. r. N=13	+.054	−.048	+.018	+.043	+.097	+.035	+.133	+.173	+.088	−.003	+.145	+.133	+.136	−.036		

TABLE IX

Coefficients of Correlation between Strength Scores

	S−1 St/Wt	S−2 C.C.	S−3 W's A.	S−4 Maze	S−5 Mün.	S−6 Can.	S−7 Cl. P.	S−8 Myst.	S−9 R. Id.	S−10 Sug.	Blood Pres.	Temp
S−1	+.*749*											
S−2	+.003	+.*833*										
S−3	+.162	−.142	+.*425*									
S−4	−.163	+.070	−.217	+.*429*								
S−5	−.201	+.222	+.124	+.159	+.*869*							
S−6	+.205	−.062	+.028	+.424	+.264	+.*800*						
S−7	+.111	+.051	+.130	+.173	−.015	+.246	+.*638*					
S−8	−.268	+.176	−.140	+.25	+.071	+.322	+.291	+.*502*				
S−9	−.103	−.083	−.043	−.187	−.101	+.261	+.006	+.126	+.*764*			
S−10	−.048	+.007	−.009	+.034	+.01	+.364	+.22	+.332	+.384	+.*567*		
Bl. P.	+.179	+.009	−.016	−.302	+.018	+.227	−.113	−.187	+.247	+.362	+.*716*	
Comp. Sc. 1.	−.275	+.231	−.083	+.432	+.343	+.552	+.38	+.47	+.292	+.46	+.058	−.129
Comp. Sc. 2				+.412	+.302	+.652	+.416	+.584	+.248	+.565		
Av. r. N = 9	−.034	+.045	−.012	+.06	+.059	+.228	+.135	+.127	+.029	+.145		

tween the 10 experiments, 30 are positive, but only 1 is 5 times its P.E. and 2 are 4 times it. When correlated with the composite score, 4 of the 10 are at least 5 times the P.E. and 1 is 4 times it. Here again the experiments in which speed and accuracy are stressed equally appear to have some degree of interrelation; also the coefficient of correlation between the test for resistance to suggestion, where the rate factor did not enter

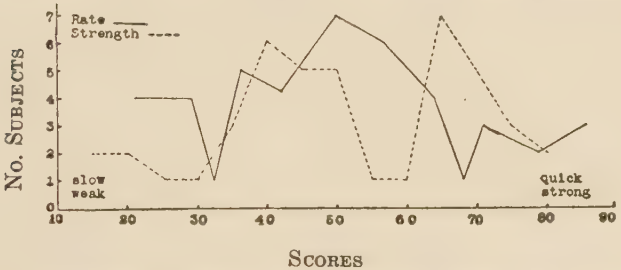


Fig. I. Frequency Curves for Composite Experimental Rate and Strength Scores. Scores by Percent Method

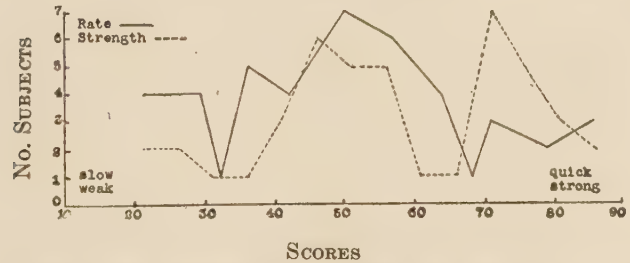
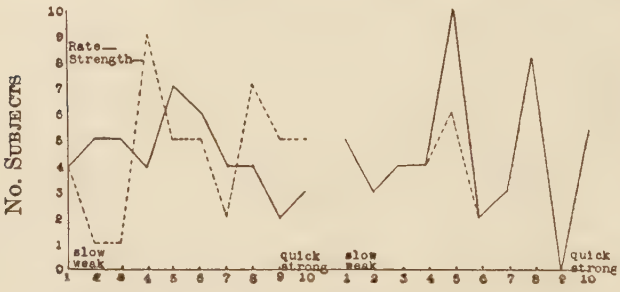


Fig. II. Frequency Curves for Composite Experimental Rate and Strength Scores with Strength Curve shifted so that the Ranges Coincide

at all, and the experiments in which quality rather than speed element predominated, suggest a closer degree of interrelation than do most of the other coefficients.

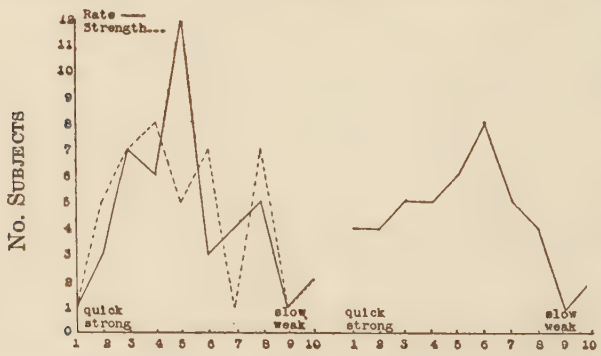
If we assume that a common rate factor and a common strength factor may exist, the question then arises whether there will be found any grouping of the *S*s with respect to these traits. Frequency curves for the composite rate, composite strength, and a combined rate and strength score are given in Figs. I, II, III, and IV. Those in Figs. I and II are plotted from the actual composite scores, whereas in all later curves the range of scores has been divided into 10 and the scores grouped accordingly. The range for the strength scores is just 6 points lower on the scale of 0—100%, therefore the distributions given in Fig. I

were plotted in Fig. II with the strength scores all shifted 6 points up on the scale, so that the position of the high and low points of the two curves might be compared. The strength curve is unquestionaby bi-modal; the rate curve can scarcely be so considered, though it does rise to a lesser extent where peaks are found in the strength curve and falls to points equally low at the lowest places in the strength curve. The agreement be-



(A) Scores (B) Scores

Fig. III. A. Composite Experimental Rate and Strength Scores (Percent Method). B. Combined Scores, Rate and Strength



(A) Scores (B) Scores

Fig. IV. A. Composite Experimental Rate and Strength Scores (Av. R. Method). B. Combined Scores, Rate and Strength

tween the curves is sufficient to suggest that a single factor might be influencing both the rate and strength aspects of the experiments. There is no evidence of this however in the coefficient of correlation for these two composite scores; it proves to be $-.0025$. Fig. V is a graphic representation of this zero correlation.

In Fig. III-A the scores plotted are the same as those in Figs. I and II. They are, however, plotted as they fall into the 10 groups into which the entire range has been divided. The nature of the curves remains unchanged. For Fig. III-B each *S*'s rate and strength scores (percent method) were added and the total plotted. Curve B is definitely bi-modal. There appears to be a quick-strong group, and another group which a study of the 10 cases falling on the fifth abscissa shows to be

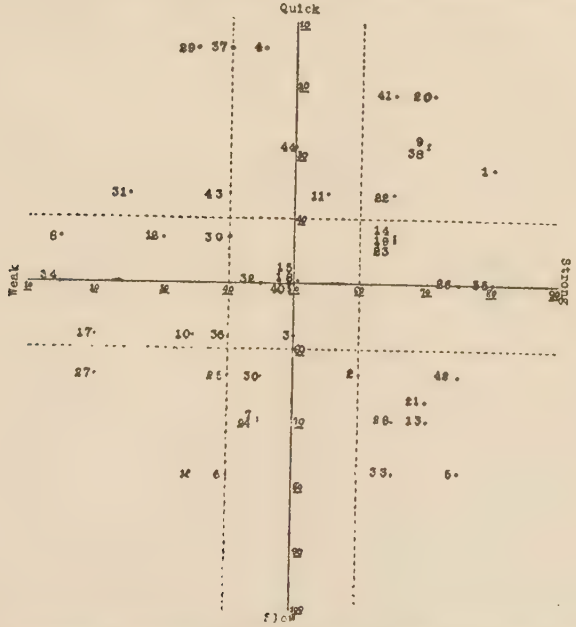


FIG. V. GRAPHIC REPRESENTATION OF THE LACK OF CORRELATION BETWEEN RATE AND STRENGTH SCORES

composed of 6 *S*s with a low rate score and a high strength score that would classify them as slow-strong and 4 cases whose scores were both close to 50%. The line dotted between abscissae 4 and 6 indicates where the slow-strong group alone would fall. There are no quick-weak *S*s in this group. There is the suggestion in this curve of a possible 3rd group of slow-weak individuals.

The scores plotted in Fig. IV, A and B, are the composite scores found by the average rank method, and are plotted as in Fig. III. The separate rate and strength curves have much the same form, but the curve for the combined scores loses its bi-modality. As has been said before, we do not consider this

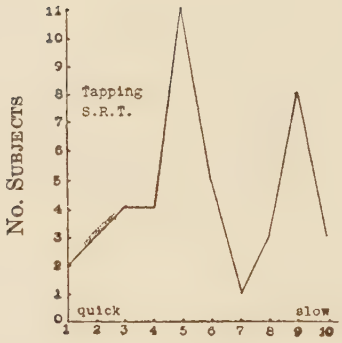


Fig. VI. Composite Score for Experiments where Rate is only Factor

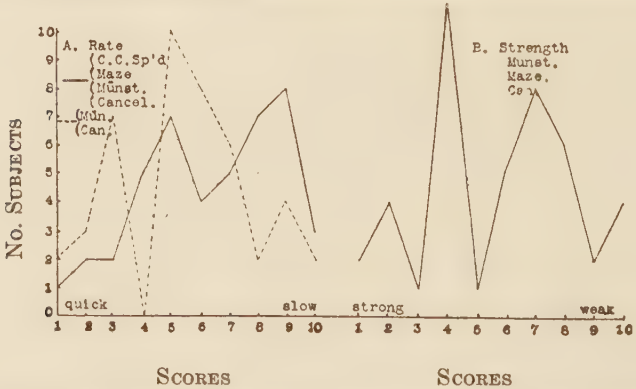


Fig. VII. Composite Scores for Experiments with Equal Stress on Speed and Accuracy

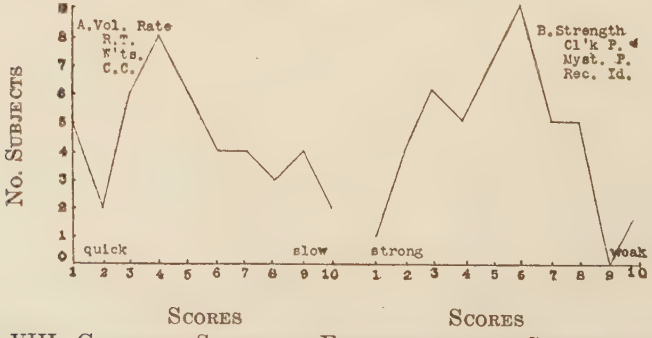


Fig. VIII. Composite Scores for Experiments with Stress either on Rate or on Quality

method of computing the composite scores as satisfactory as the percentage method, since undue weight may be given to an extreme score opposed to the general tendency of *S*'s performances, when this reversal may have been caused by a sudden shift in attitude due to extraneous factors.

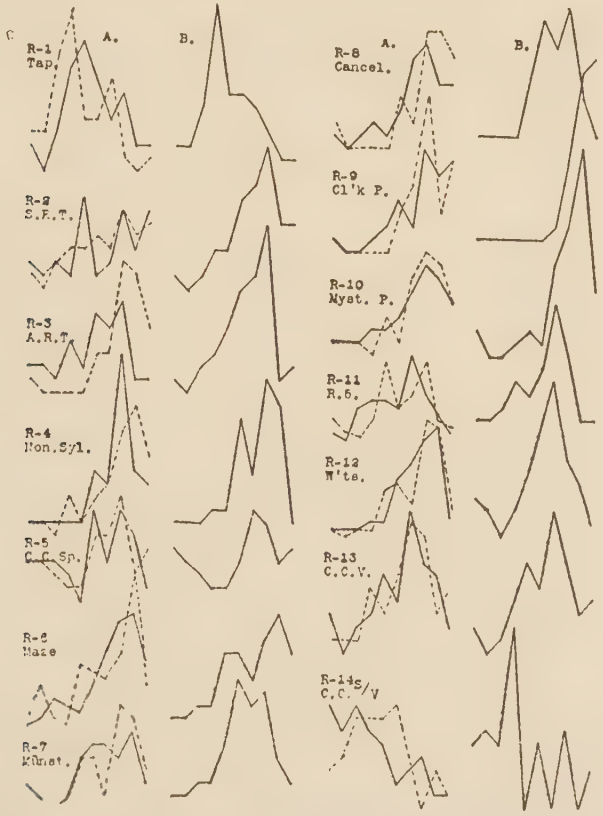


FIG. IX. FREQUENCY CURVES FOR THREE SETS OF SCORES FOR EVERY RATE EXPERIMENT

We next broke up the composite scores into scores for the various types of rate and strength experiments. The average rank method was used to obtain these scores, since 4 was the largest number of experiments in any group. These frequency curves are given in Figs. VI, VII, A and B, and VIII, A and B. The curves both for rate and for strength in the experiments requiring an equal effort for speed and accuracy are unquestionably bi-modal; as is also the curve for the average score for

tapping and simple reaction time. Bi-modality is much less marked in the curve for the experiments in which stress was laid on speed and is not present in the one for those where emphasis was on quality.

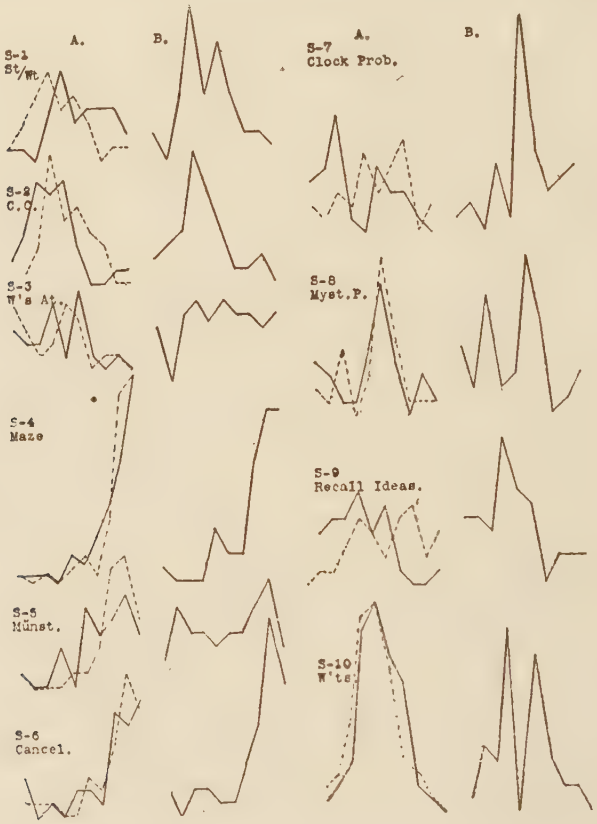
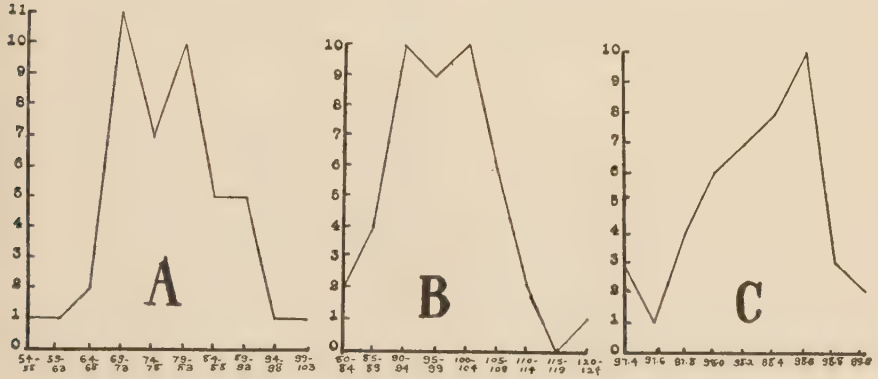


FIG. X. FREQUENCY CURVES FOR THREE SETS OF SCORES FOR EVERY STRENGTH EXPERIMENT

The search for bi-modality was carried further. We had three sets of scores for every experiment, the two sets from which the coefficients of reliability were computed and the final data. Curves were plotted for these 72 sets of scores to determine whether any experiment consistently yielded this form of curve. These curves are to be found in Figs. IX and X. The double curves in the A columns are from the scores obtained for the coefficients and those in the B columns from the final data. Of the 14 sets of rate curves the ones for *R-5* (carbon-copy speeded

time) and for *R-14* (ratio of carbon-copy speeded to carbon-copy voluntary time) are the only ones which give even the slightest indication of bi-modality in all 3 curves for any experiment. Inspection of the 10 sets of strength curves yield similar results. The set for *S-8* (mystery problem) is the only one where more than one mode can be considered to exist in all of the curves.

The results which have been studied in the form of coefficients of correlation and in frequency curves cannot be said to show definitely that there is neither a common rate nor a common strength factor. On the other hand they do not by any means indicate conclusively that such factors exist; they suggest rather that there may be such a possibility.



PULSE RATE BLOOD PRESSURE TEMPERATURE

FIG. XI. FREQUENCY CURVES FOR PHYSIOLOGICAL MEASURES: A. PULSE RATE; B. BLOOD PRESSURE; C. TEMPERATURE

Curves for the physiological measures are given in Fig. XI. Neither the blood pressure nor the temperature curves give any indication of types. The curve for the pulse rate measures may be considered bi-modal and suggestive of a group of individuals with an unusually slow pulse in addition to the group whose rate agrees with what is accepted as normal. When the pulse rate for the cases scoring highest or lowest in the rate experiments is studied an unexpected relation appears to be present in spite of a coefficient of correlation of .098. Seven of the 8 *S*s receiving the highest experimental rate scores have an average pulse rate below 79, which is the average for all 44 cases, and 24 of the 31 measures for these *S*s are below the average, while 5 of the 8 with the lowest rate scores have an average pulse rate above 79 and 21 of the 32 measures are above it. It may be that the high pulse rate in these cases is an indication of an excited

condition which interfered with success in the experiments, especially with respect to speed. On this point it is interesting to note that the coefficient of correlation between pulse rate and rate of emotional arousal is .303; this is not high, but it is suggestive. The only *S* (BeC) who has a high pulse rate with a high experimental rate score does fall down in her strength score. Among the 5 with high pulse rate but low experimental rate scores, 4 also have low strength scores; of the 3 who have low pulse rates and low experimental rate scores, all did well in the matter of strength. A similar study with respect to blood pressure and temperature measures, and experimental scores, does not disclose any hint of a connection between them.

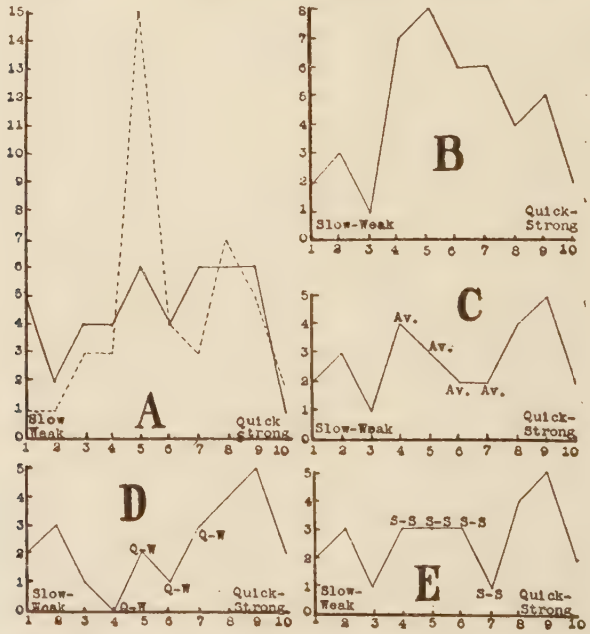


FIG. XII. QUESTIONARY: COMPOSITE RATE AND STRENGTH SCORES

A. Rate and Strength scores given separately.
B. Rate and Strength scores combined.
C. Average cases.
D. Quick-Weak cases.
E. Slow-Strong cases.

C, D, and E are plotted from the scores used in B. In them the modal group has been broken up into Quick-Weak and Slow-Strong cases and cases average in both traits. The rest of the curve remains unchanged. The *abscissae* represent the *scores*, the *ordinates* the *number of subjects*.

The frequency distribution of the composite rate scores and composite strength scores obtained from the questionaries is given in Fig. XII. The curves in part A are for these two sets of scores. Again the strength curve is more definitely bi-modal than the rate curve; the latter has 3 peaks of nearly equal height, suggesting a slow, an average, and a quick group. When the two composite scores are combined we have curve B which indicates a small slow-weak group, a larger quick-strong group, and a still larger middle group composed of the quick-weak and slow-strong cases and those average in both traits. The cases in this center group were sorted out and the next 3 curves plotted to show their distribution. In curve C only the cases average in both traits appear, in D the quick-weak, and in E the slow-strong cases; all are to some extent tri-modal. The coefficient of correlation between the composite rate and strength scores is $-.107$.

The degree of correlation existing between the other tendencies investigated by the questionary is as follows:

Steadiness and Composite Strength Score (Quest.)	$= -.557$
Activity and Composite Rate Score (Quest.)	$= +.626$
Emotional Intensity and Composite Strength Score (Quest.)	$= +.192$
Emotional Arousal (Speed) and Composite Rate (Quest.)	$= +.788$

When we come to investigate what relation exists between the experimental, physiological, and questionary results, we find that for the most part the correlations are negligible. See Tables VIII and IX for most of the correlations between physiological measures and experimental scores. Others, and those between the physiological measures and questionary results, are given below:

Blood Pressure Range on First Day and Comp. Exp. Rate Score	$= +.275$
Blood Pressure Range on First Day and Comp. Quest. Rate Score	$= +.164$
Pulse Rate and Comp. Quest. Rate score	$= -.048$
Average Blood Pressure Scores and Comp. Quest. Strength Score	$= +.019$
Pulse Rate and Rate of Emotional Arousal	$= +.303$

To make a statistical study of the relation between questionary measures of rate and strength and the experimental scores, we wished to obtain as many correlation coefficients as possible between measurements secured by these two methods. We selected first the combinations most likely to give a correlation, next we used the experiments with the highest coefficients of reliability. When 232 of the possible 555 combinations had been plotted and partially worked out and 54 of the 119 rate and 35 of the strength coefficients had been found, the results

were such that it did not seem profitable to go further with this undertaking. The extreme coefficients for the rate and strength groups as far as they were completed are as follows:

Quest. 2. (Speaks rapidly) and *R-11* (Reading Time) = +.355
Quest. 22 (Quick to note changes) and *R-10* (Myst. Prob.) = +.280
Quest. 3 (Works rapidly) and *R-1* (Tapping) = +.256
Quest. 13 (Quick in mental processes) and *R-3* (Assoc. R. T.) = +.228
Quest. 13 (Quick in mental processes) and *R-9* (Clock Prob.) = −.250
Quest. 2 (Speaks rapidly) and *R-9* (Clock Prob.) = −.245
Quest. 2 (Speaks rapidly) and *R-3* (Assoc. R. T.) = −.242
Quest. 47 (Acts impulsively) and *R-6* (Maze) = −.210
Quest. 56 (Self-confident) and Comp. Exp. Strength Score = +.320
Quest. 37 (Intense likes and dislikes) and *S-2* (Carbon-Copy) = +.262
Quest. 10 (Dominates conversation) and *S-2* (Carbon-Copy) = +.262
Quest. 4 (Works accurately) and *S-7* (Clock Prob.) = +.262
Quest. 41 (Calm in emergencies) and *S-7* (Clock Prob.) = +.252
Quest. 46 (Perseveres in difficulties) and Comp. Exp. Strength Score = +.236
Quest. 23 (Notices details) and Comp. Exp. Strength Score = −.223
Quest. 34 (Exerts self-control in emotion) and Comp. Exp. Strength Score = −.219
Quest. 15 (Appreciates humorous sayings) and Comp. Exp. Strength Score = −.207

Other coefficients found between composite experimental or questionary scores and scores for groups of experiments or groups of questions are given below:

Comp. Rate Score, 14 exps., and Comp. Quest. Rate Score = +.061
Comp. Rate Score, 8 exps., and Comp. Quest. Rate Score = +.052
Comp. Strength, 10 exps., and Comp. Quest. Strength Score = +.022
Comp. Strength Score, 7 exps., and Comp. Quest. Strength Score. = +.041
Comp. Quest. Rate Score and Tapping (speed only) = +.086
Comp. Quest. Rate Score and Av. R. in Tapping and Simple Reaction Time (speed only) = +.079
Comp. Quest. Rate Score and Av. R. in Carbon-Copy Sp'd, Maze, Münst., and Cancel. (speed and accuracy) = −.114
Comp. Quest. Rate Score and Av. R. in Münst. and Cancel. (speed and accuracy) = −.040
Comp. Quest. Rate Score and Av. R. in Reading Time, Judg.. W'ts. Carbon-Copy Vol. (voluntary rate) = +.227
Steadiness and Comp. Exp. Strength Score = +.077
Activity Score and Comp. Exp. Rate Score = +.092
Emotional Intensity and Comp. Exp. Strength Score = −.207
Emotional Intensity and Comp. Exp. Rate Score = −.068

The negligible degree of correlation, in the case of both rate and strength, between experimental and questionary scores may be expressed differently. We find that 7 cases receive the same classification for both traits by the experimental and questionary scores; in 12 instances the classification is not changed to the opposite one for either trait, but one or more of the four scores are just average; in 18 cases either the rate or strength classification by one set of results is opposed to the classification by the other set, and in 7 cases both rate and strength classifications by one set are contradicted by the other.

The statistical treatment of the interrelation of results was supplemented by a study of the two extreme cases from each of the four groups, based on the experimental score. The cases are numbers 1 and 20 from the quick-strong group, 29 and 37 from the quick-weak, 5 and 23 from the slow-strong, and 16 and 27 from the slow-weak group. (See Fig. V for the grouping of the cases.) In this study we sometimes find a striking similarity between experimental results and parts of the questionary results, but equally often a marked dissimilarity, and at no time is the agreement perfect. The fact that of the 16 cases rated as definitely weak by the questionary scores only 4 are also rated slow, and that only 4 of the 21 whose scores classify them as slow are also considered definitely weak, suggests that those who filled out the questionaries may have been loath to give low ratings for both types of traits, but tended rather to balance them. Four of the case studies, one from every group, are given in illustration:

Case 1. A. An experimental rate score of 68, an experimental strength score of 80, a questionary rate score of 89 and a questionary strength score of 69 give this *S* a classification of quick-strong on a basis of both sets of results. In Questionary C she receives a rank above the average rank of the 44 cases in each of the rate traits except for rate of talking, which is just average. In the 29 traits taken as suggestive of strength her rank is above the average in 20 instances; in the other 9 she is very close to the average except in the control of her emotions and calmness in emergencies. She appears to be quick physically, mentally, and emotionally, to be forceful, energetic, and dominating, to be a steady worker, resolute and fixed in her opinions, but to be uneven in mood or disposition and to have changing antipathies and sympathies. She is ranked approximately average for optimism but below average for good-naturedness and slightly below in tendency to praise. This is suggestive of the unpleasant feeling tone which Ebbinghaus, Kreibig, and Meumann[51] among others attribute to this temperament. Another indication of this feeling quality is found when we consider the values assigned to the 35 activities listed in Questionary D; these values indicate the amount of pleasure derived from participation in the activities. To 5 activities she assigns the value which is average for the 44 *S*s, to 12 she assigns a value above the average, but one below average to 18 of them. From 11 of these 18 she derives no pleasure at all; even displeasure from 8 of them. Contrary to what we should expect of persons in this group, the data here for the most part show a disinclination for activities involving much physical exercise, also distaste for competitive games either physical or mental. This may be partly explained by a record at the University Health Department which states that her home physician advised against gymnasium because of a heart condition; this weakness may be a factor in her dislike of exercise. However, we find that in Questionary C her total score for the tendency to be active is +4, where the range of scores for all 44 cases is −6 to +6 and both average and median is 0. When we consider the next part of Questionary D, we find that *S* reports that she is a sound sleeper, that the number of nights during which she dreams is 25% where the average for all *S*s is 18.6%, but that she dreams only 5% of the average night while the average for the 36 *S*s who report any dreaming is 34.5%. She estimates her pleasant dreams at 50%

[51] W. Stern, *op cit.*, 483 f.

and her unpleasant at 25%. In our search further for evidence of nervousness we find little; she states that she very seldom bites her nails now although she often did between the ages of 4 and 12 years, that she seldom rocks while in a rocking chair, but does much of the time swing her free foot when sitting with her knees crossed. There appears in general to be fairly good agreement throughout the data on this case.

Case 29. M. Experimental scores of 86 and 35 classify *M* as quick-weak; questionary scores of 64 and 86 change the classification to quick-strong. Though the ratings are not usually very much above the average for the trait, she is considered quick in most respects; exceptions are slowness in speech and slowness to anger and impatience. When we consider the strength questions, we find that the only weak points stressed are the ease with which she is persuaded and the mild intensity of her anger. Most of the traits in this group are rated unusually high and picture her as an accurate and steady worker, energetic, industrious, resolute, self-controlled, calm in emergencies, self-confident, enthusiastic, and given to intense likes and dislikes. Optimism and a pleasant feeling tone are assigned to the quick-weak temperament, and this *S* is rated high in them. This characteristic is suggested also by her record in Questionary D, for she gives a value above average to 25 and one below average to only 4 of the activities; she derives the greatest amount of pleasure from 16 activities, and no pleasure at all, or in this case displeasure, from only one. She enjoys both physical and mental activity. *M* reports that she is a light sleeper, dreams about 75% of every night and that half her dreams are pleasant, the other half being equally divided between neutral and unpleasant. Though she often bites her nails she seldom rocks and never swings her foot. Her record at the University Health Department contains a report of "nervous trouble, cried, little things bothered." In the daily questionary she made notes on the first and fourth days that she was "nervous," but her low strength scores were not confined to those days. The data on this case are throughout not in harmony; the ratings on the strength traits are opposed to the evidence gained from all other sources.

Case 33. P. Slow-strong is the classification earned by experimental scores of 21 and 65 and slow-average by questionary scores of 29 and 48. Only in talking and working is *P* considered really quick. She is thought slow in her movements and mental processes, slow to anger and in the arousal of other emotions. The outstanding strength values indicate that she is a steady worker, resolute and courageous, and exerts great self-control in showing her emotions, but that she is inclined to be mentally indolent and inaccurate in her conclusions and opinions, to become only mildly angry and to be indifferent toward people and things. Instead of the tendency toward pessimism and unpleasant feeling tone expected of persons in this group, we find here an inclination to be cheerful and happy, to praise and be optimistic, although in contradiction to this she is also rated as habitually anxious and worried. The values assigned by her to the activities do not give evidence of a definite tendency in either direction; she gives a value above average to 15 but one below average to 13, derives the greatest pleasure from 8, but no pleasure at all, and in one instance displeasure, from 7. She claims to be a sound sleeper, to dream during few nights (5%), and not to any great extent (10%) at those times. Her dreams are chiefly neutral and only 5% are unpleasant. She also seems to be troubled little by nervous habits, for she never bites her nails, though she often did when about 10 years of age; she rocks little of the time that she is seated in a rocking chair, and seldom swings her free foot when her knees are crossed. Again, the data from various sources do not run parallel throughout.

Case 16. F. An experimental rate score of 21 and experimental strength score of 35 classify *F* as slow-weak. Her questionary rate score of 7 agrees with this; her questionary strength score of 48 places her with the average group. Most of the values for the rate traits are well below the average and those for the strength traits close to it. She is considered slow in every way, careful, accurate and steady in work; to feel losses for a long time, to be resolute, perserveing and self-controlled, but to lack force and energy, to have changing likes and dislikes, to be easily persuaded and wanting in self-confidence. We should expect persons in this group to give evidence of pleasant feeling tone. This is true in her case for she is described as good natured, cheerful and happy though inclined to alternate in mood; as given to praising and laughter, and as optimistic. There is further evidence of this tendency in the values assigned to the activities, for she gives a better than average score to 15 and a score below average to only 7. To 11 of the 15 she gives the highest possible score, while she claims to derive no pleasure or even displeasure from only 3. Activities involving either physical or mental exertion are scored high, although her total score for activity in Questionary C is −4; the latter is more in keeping with what one would expect of a person classified as weak and slow. *F* reports that she sleeps soundly, dreams during about one night in 4 and only about 25% of those nights; 50% of her dreams are pleasant and 25% neutral. She apparently does not have any of the common habits suggestive of nervousness, for she very seldom bites her nails, seldom rocks, and very seldom swings her foot. The conservative judgments concerning most of the strength traits weaken an otherwise close agreement between the various sets of data.

IV. Conclusions

(1) There is a slight indication of the possibility of a common rate factor in the experiments in which speed and accuracy were equally stressed and in those where the rate was voluntary.

(2) A common strength factor is suggested to a slight degree by the results of the experiments in which speed and accuracy were stressed equally and in those where the emphasis was on the quality of the performance.

(3) The 44 cases do not fall into four sharply defined groups on a basis of the experimental scores, but at the same time there do seem to be a few cases who are consistently quick and are either consistently strong or consistently weak, and others consistently slow, who are fairly consistent in their strength or weakness.

(4) The physical measures give practically no evidence of types and do not in general appear to be consistent with the experimental results. An exception to this is the negative relationship between the extreme experimental rate scores, both high and low, and pulse rate. We suggested above that this may be indicative of an excited state which interfered with the performance of the experiments.

(5) The questionary rate and strength scores also do not divide the *Ss* into distinct groups, but again, as with the experi-

mental scores, there appear to be some cases who are consistent in the matter of both rate and strength.

(6) The agreement between the experimental and questionary classifications is slight enough to be accounted for by chance.

(7) In general we must conclude that although there may be general rate and strength factors as we have defined them, still, in so far as the questionaries yield evidence concerning temperament, these factors do not seem to be related to temperament.